Appropriate temporal resolution of precipitation data for discharge modelling in pre-alpine catchments

Sikorska, Anna E ; Seibert, Jan

Abstract: Precipitation time series with high temporal resolution are desired for hydrological modelling and flood studies. Yet the choice of an appropriate resolution is not straightforward because the use of too high a temporal resolution increases the data requirements, computational costs and, presumably, associated uncertainty, while performance improvement may be indiscernible. In this study, the effect of averaging hourly precipitation on model performance and associated uncertainty is investigated using two data sources: station network precipitation (SNP) and radar-based precipitation (RBP). From these datasets, time series of different temporal resolutions were generated, and runoff was simulated for 13 pre-alpine catchments with a bucket-type model. Our results revealed that different temporal resolutions were required for an acceptable model performance depending on the catchment size and data source. These were 1–12 h for small (16–59 km2), 3-21 h for medium (60–200 km2), and 24 h for large (200–939 km2) catchments.

DOI: https://doi.org/10.1080/02626667.2017.1410279

Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: http://www.tandfonline.com/loi/thsj20

Appropriate temporal resolution of precipitation data for discharge modelling in pre-alpine catchments

Anna E. Sikorska & Jan Seibert

To cite this article: Anna E. Sikorska & Jan Seibert (2017): Appropriate temporal resolution of precipitation data for discharge modelling in pre-alpine catchments, Hydrological Sciences Journal, DOI: 10.1080/02626667.2017.1410279

To link to this article: https://doi.org/10.1080/02626667.2017.1410279

Publisher: Taylor & Francis & IAHS

Journal: *Hydrological Sciences Journal*

DOI: 10.1080/02626667.2017.1410279

Appropriate temporal resolution of precipitation data for discharge modelling in pre-alpine catchments

Anna E. Sikorska[1,2,*] and Jan Seibert[1,3]

[1] *Department of Geography, University of Zurich, Zurich, Switzerland*

[2] *Department of Hydraulic Engineering, Warsaw University of Life Sciences - SGGW, Warsaw, Poland*

[3] *Department of Earth Sciences, Uppsala University, Uppsala, Sweden*

**Corresponding author. Tel. +41 4463 56532; e-mail: as@annasikorska.eu*

Abstract

Precipitation time series with high temporal resolution are desired for hydrological modelling, especially for flood studies. Yet, the choice of an appropriate temporal resolution is not straightforward because the use of time series with too a high temporal resolution increases the data requirements, computational costs and presumably associated uncertainty, while performance improvement may be indiscernible. In this study, the effect of averaging hourly precipitation on model performance and associated uncertainty is investigated using two data sources: station network precipitation (SNP) and radar-based precipitation (RBP). From these datasets, time series of different temporal resolutions were generated, and runoff was simulated for 13 pre-alpine catchments with a bucket-type model. Our results revealed that different temporal resolutions were required for an acceptable model performance depending on the catchment size and data source. These were 1–12 h for small (16–59 km^2), 3, 12 and 21 h for medium (60–200 km^2), and 24 h for large (200–939 km^2) catchments.

Keywords radar-based precipitation; station network precipitation; averaging length; uncertainty; Bayesian methods; bucket-type model

1 Introduction

A high temporal precipitation resolution is generally desirable for hydrological modelling and predictions (Wang *et al.* 2009), especially when the focus lies on floods. However, if the

temporal data resolution is higher than the temporal scale of the runoff formation processes, the higher resolution may not necessarily lead to improved model performance (Bastola and Murphy 2013), as information contained in data is limited (Kuczera *et al.* 2010, Sikorska and Seibert 2016), but might put too high requirements on data availability.

Requirements of a high temporal resolution (sub-daily or finer) of precipitation data usually cannot be met at poorly gauged or remote sites (Reynolds *et al.* 2015), such as mesoscale pre-alpine catchments (area in the range of roughly 15–1000 km^2 in the analysed case). For these type of sites, high resolution data is limited to a few recent years at best, whereas daily records are usually available for a longer period. For example, hourly stations in Switzerland have been regularly operated since the 1990s, while daily information is often available back to the 1930s (source: MeteoSwiss). Moreover, not only the record length but also the number of available stations is usually much smaller for hourly than for daily data (Koutsoyiannis and Onof 2001), and often limited to urbanized sites. This has its reasons in lower costs, time demands and ease of collection (Pui *et al.* 2012). Thus, in many practical applications, the data availability remains the major factor for the choice of data resolution for model calibration (Kavetski *et al.* 2011), leading often to a lower resolution chosen than actually required (Aronica *et al.* 2005).

In contrast, using an unnecessarily high resolution is generally linked to increased computational costs, as the amount of data points over the same period multiplies. More data points combined with long time series may become critical for uncertainty analysis based on Monte Carlo techniques with thousands of model runs required (Sikorska *et al.* 2015c). Hence, some form of data aggregation may be indeed useful to reduce computational demands while keeping information content from high-resolution data. The simplest aggregation approach relies on the concept of averaging the precipitation totals over desired sub-periods but has two major drawbacks for hydrological models. First, averaging precipitation totals causes the loss of information contained in high-resolution data. This is of a major concern for floods in catchments with a fast response time (Wetterhall *et al.* 2011). Second, aggregating yields an additional uncertainty due to the averaging method itself and representativeness of such estimated precipitation inputs at the temporal and spatial catchment scale. This uncertainty adds to other uncertainty sources that are parametric uncertainty of a hydrological model, model structural errors, and errors in runoff data (Thyer *et al.* 2009, Renard *et al.* 2011, Sikorska *et al.* 2015a).The contribution of this averaging uncertainty to the total uncertainty is unclear because averaging may decrease the relevance of precipitation measurement uncertainty which, being important at recorded high resolutions, may become smoothed at greater temporal scales.

As the temporal resolution of precipitation data is essential for hydrological modelling, a number of studies have investigated its effect on parameters of hydrological models (Booij 2002, Hearman and Hinz 2007, Littlewood and Croke 2008, Patil and Stieglitz 2015), or the selection of an optimal resolution (Ochoa-Rodriguez *et al.* 2015, Wetterhall *et al.* 2011). Yet, the effect of averaging precipitation totals over temporal compartments on the runoff model performance and associated uncertainty has gained only little attention until now. Earlier research has mostly focused on parametric uncertainty of runoff models using an informal uncertainty framework (Aronica *et al.* 2005) neglecting model structural errors, or a formal uncertainty framework (Kavetski *et al.* 2011) but with a simplistic error description assuming *homoscedastic*, i.e. independently and normally distributed model residuals. It has been shown however that an inadequate treatment of structural errors will generally bias estimates (Renard *et al.* 2011). Moreover, most of previous studies focused on urban catchments while pre-alpine regions remained of less interest. Due to their complex terrain and possibly higher errors in estimation of the catchment precipitation, these catchments require special consideration (Joss and Lee 1995, Sikorska and Seibert 2016).

The aforementioned issues make clear that an objective way to determine an optimal precipitation temporal resolution in pre-alpine catchments and/or the suitable degree of precipitation data averaging with the ability for reliable uncertainty estimates is important. Therefore, the aims of this paper were (i) to examine the value of different degrees of precipitation data aggregation using a more reliable model error description, i.e. assuming *heteroscedastic* (correlated and non-normally distributed) residuals and a simple bucket-type model; (ii) to guide an optimal selection of precipitation temporal resolution in the range of one hour to 48 hours for runoff modelling in pre-alpine mesoscale regions using Bayesian statistics. These aims were investigated with two different precipitation data sources: station network precipitation (SNP) and radar-based precipitation (RBP), which are particularly beneficial for hydrological modelling in pre-alpine catchments.

In this work, we focus solely on the effect of increasing the precipitation temporal averaging length on the runoff model performance and associated uncertainty, while other issues related to a simulation time step of the hydrological model or a temporal dependency of model parameters are not considered.

2 Catchments, precipitation datasets, and catchment grouping

Our study is based on 13 mesoscale Swiss catchments of different physiographic characteristics with an average altitude ranging from 605 to 1917 m a.s.l., and an areal range of 16.2– 939 km^2 (Fig.1 and Table 1). The catchments were selected to cover a range of sizes and are not (or only slightly) influenced by human activities.

For these catchments, detailed data of temperature, precipitation, and runoff were available for the period 2005–2010 with an hourly resolution, as well as average daily temperature and evapotranspiration estimates from long-term observations (source: MeteoSwiss and FOEN). Precipitation information for each catchment was acquired from two different sources: station network precipitation (SNP) and radar-based precipitation (RBP). The SNP was generated from the automatic national gauging network comprised of about 75 Swiss ground stations (tipping-bucket type gauges) with a 10-min recording resolution. The mean areal catchment precipitation from SNP data source was generated by averaging the precipitation totals recorded at the respective ground stations located within and near each catchment, using a commonly applied (also in mountainous regions) Thiessen method (Fig. 1), and summed up to hourly resolutions. The RBP is a precipitation product provided by MeteoSwiss generated using records from three Swiss weather radars adjusted to precipitation amounts with high quality daily ground stations (Germann et al., 2006). This dataset has been provided for the whole of Switzerland at an hourly resolution and with a spatial grid size of 1 km^2 but is available only for a limited number of years. A post-process correction for bias is applied to these estimates prior to usage which allows significantly reducing the measurement error. RBP was spatialized to each catchment area as an average over grids falling within the catchment borders.

The catchments were grouped according to area: those with a catchment area, A, smaller than 60 km^2 were attributed to small catchments, those with an area in the range of 60–200 km^2 to medium size catchments, and those with an area equal to or larger than 200 km^2 to large catchments (Fig.1).

3 Methods

3.1 Assumptions

Our work is based on the hypothesis that precipitation amounts can be assumed for runoff modelling as uniformly distributed in time over the given (averaging) window length α (Section 3.2). This hypothesis is supported by the assumption that (a) averaged data can be aggregated to minimize the amount of data points and in this way advance model simulations, and (b) as an inverse approach, sparse data could be disaggregated into finer resolutions and thus the timely variation of precipitation at finer scales cannot be known. We test this hypothesis with 17 calibration schemes of different averaging window lengths α and two different precipitation datasets (SNP and RBP, Section 2). In this way, we are able to identify to which degree the loss of information contained in finer data becomes (un-)important for hydrological modelling.

We further assume that the spatial distribution of precipitation data at finer scales corresponds to this observed at a sparser scale, which can be applied to the mean catchment precipitation as in bucket-type models (e.g. a HBV model used here, Section 3.3.1). This assumption might be violated if precipitation amounts were spatially re-distributed, e.g. when using a distributed runoff model.

In our work we also made an explicit division between only two uncertainty sources, i.e. parametric and a remaining uncertainty term represented with the error model (Section 3.3.2). The latter lumps uncertainty due to model structural limitations, model input (precipitation, temperature, evaporation) and model output (runoff calibration data), as well as all other remaining uncertainty sources. In this work we however assume the relative differences observed in the error term to be dominated by precipitation uncertainty. This can be assumed because only precipitation inputs (and specifically their temporal distribution) are varied between different schemes. In addition, by running the runoff model always at hourly time steps, other effects such as temporal parameter dependency or numerical issues, could be excluded.

3.2 Approach to temporal averaging of precipitation datasets

The available hourly precipitation datasets of both sources, i.e. station network (SNP) and radar-based (RBP), were treated by using a simple temporal averaging approach, where the only parameter is the averaging window length α. In this approach, we assume the uniform distribution for precipitation amounts over the averaging window. Hence, this approach is somehow similar to a running mean concept when data records are averaged over a defined window length.

The averaging approach has the following steps (see also Fig. 2):

Step 1: Choose an averaging window length α in the range of 1 to 48 h, where averaging over 1 h equals the original hourly precipitation datasets (i.e. no aggregation).

Step 2: Aggregate the hourly precipitation dataset over the chosen averaging window α, resulting in cumulated precipitation amounts over each window of the length equal to α.

Step 3: Assuming uniform distribution of precipitation amounts, uniformly redistribute the aggregated precipitation amounts over α unique compartments each corresponding to an hourly time interval.

Step 4: Repeat steps 1–3 for a newly selected window length.

Seventeen different window lengths in the range of 1 to 48 h were tested, i.e. from 1 to 6 h, every hour (i.e. 1, 2, 3, ..., 6 h); from 6 to 12 h, every two hours; from 12 to 24 h, every three

hours; and from 24 to 48 h, every six hours. The results of these 17 precipitation averaging schemes were used as input data for the runoff model (Section 3.3). The 1-h data were used as a reference for comparing model results. Note that the averaging always started at the first hour of the day (1–24 h) so that the precipitation totals over a day remain consistent between different averaging window lengths.

3.3 Stochastic description of the runoff model

3.3.1 Semi-distributed runoff model (HBV)

Runoff was modelled with the HBV model (Bergström, 1992), which has been successfully used in mountainous (Blöschl et al. 2007; Breinl 2016; Jost et al. 2012) and particularly Swiss catchments (e.g. Griessinger et al. 2016; Nans et al. 2014; Sikorska *et al.* 2015b; Staudinger et al. 2015; Sikorska and Seibert 2016). The HBV model is a semi-distributed bucket-type model, where precipitation and temperature inputs are distributed within the catchment along predefined elevation zones using a constant altitude-dependent correction factor. The HBV model consists of four main routines which represent snow processes, soil moisture, groundwater and response, and streamflow routing in the channel (for model parameters see Table 2). In this study the version HBV light (Seibert 1997, Seibert and Vis 2012) was used. Input variables of the HBV model are mean air temperature and areal precipitation, and the output is the runoff at the catchment outlet. Additionally, the model requires long-term averages of seasonally varying daily estimates of potential evaporation, which are corrected according to temperature anomalies estimated from the current temperature. The actual evapotranspiration is computed from the soil moisture.

To ensure that the simulation time step does not impact results, the HBV model was always run with the same hourly time step for all precipitation aggregation schemes Temperature data were not averaged but were always used as recorded.

3.3.2 Model errors

If the runoff model (here HBV from Section 3.3.1) could nearly perfectly reproduce the patterns between observed precipitation–runoff data, the uncertainty of the runoff model simulations could be represented with its parametric uncertainty only, meaning that the structural error of the runoff model was irrelevantly small. Yet, because of a simplified approximation of the runoff process, the runoff model is unlikely to truly reproduce observed data (Theyer *et al.* 2009). Acknowledging this fact, one may want to explicitly represent these model errors with an additional error term, which accounts for all remaining errors in runoff modelling not strictly related to model parameters.

These errors can be assumed to be *homoscedastic* if they have a constant variance over time. Thus, the sum of the runoff model and the error term can be represented by a normal distribution with an expected value equal to the output of the runoff model and an unknown (small) standard deviation. Although assuming homoscedastic model residuals greatly simplifies mathematical and computational treatment, misuse of this very strong assumption without justification will usually lead to overestimating the model goodness of fit and bias simulations (Renard *et al.* 2011), and should therefore be avoided. Indeed, errors of runoff models have been reported to be strongly correlated and *heteroscedastic* (having a changing variance over time) (Yang *et al.* 2007).

In this work we represented such correlated heteroscedastic errors with a two-parametric auto-correlated error term, as suggested by Yang *et al.* (2007). The meaning of two parameters is that σ describes the correlation strength and τ determines the correlation length among subsequent errors. To simplify model description and calculations, the error model is selected in such a way so that its (transferred) errors can still be modelled as normally distributed with a zero mean and a standard deviation σ_N computed using σ and τ at each time step t_i as:

$$\sigma_{N_{t_i}} = \sigma \cdot \sqrt{1 - \exp\left(-2 \cdot \left(|t_i - t_{i-1}|/\tau\right)\right)} \quad (1)$$

where t_{i-1} and t_i correspond to the subsequent time steps. Note that if τ goes to 0 or if $|t_i - t_{i-1}|$ is large in comparison to τ, σ_N goes to σ, meaning that the error autocorrelation becomes irrelevant.

3.3.3 Final form of the stochastic runoff model

The final form of the stochastic runoff model is simply an addition of the output from the runoff model (Section 3.3.1) and the error model (Section 3.3.2) in the transformed space, to which the normal distribution can be now assigned, with a mean equal to the output of the runoff model and a changing variance equal to $\sigma^2 \cdot \exp(-|t_i - t_{i-1}|/\tau)$.

3.4 Bayesian calibration of the stochastic runoff model

3.4.1 Bayesian inference on model parameters

According to Bayes' rule (Gelman *et al.* 2013), *a priori* defined distribution on model parameters $p(\boldsymbol{\theta})$, i.e. without using data, can be updated with, and thus conditioned on, information contained in the observation data $\boldsymbol{y}^o$ to the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}^o)$:

$$p(\boldsymbol{\theta}, \sigma, \tau|\boldsymbol{y}^o) = \frac{p(\boldsymbol{\theta},\sigma,\tau)\cdot p(\boldsymbol{y}^o|\boldsymbol{\theta},\sigma,\tau)}{\int\int\int p(\boldsymbol{\theta},\sigma,\tau)\cdot p(\boldsymbol{y}^o|\boldsymbol{\theta},\sigma,\tau)d\theta d\sigma d\tau} \quad (2)$$

where the bold font indicates a vector. This updating is executed through evaluating the likelihood function, $p(\boldsymbol{y}^o|\boldsymbol{\theta},\sigma,\tau)$, which simply measures the probability (likelihood) that $\boldsymbol{y}^o$ could have been generated with the given model, its inputs and the parameter vector sampled from $p(\boldsymbol{\theta})$. The stochastic runoff model (Section 3.3) was thus Bayesian calibrated in each catchment with 17 different averaged precipitation datasets (Section 3.2) for both data sources, i.e. SNP and RBP, always using the hourly runoff data for comparison with model simulations. This yields 34 calibration schemes per catchment and the same amount of parameter posteriors.

3.4.2 Conditional likelihood function

Because it is numerically challenging to deal with a model variance that is not constant over time, to explore the information in $\boldsymbol{y}^o$ and thus to infer model parameters $\boldsymbol{\theta}$, σ, τ, we used a likelihood function that combines the auto-correlated error model with a Box-Cox transformation (Box and Cox 1988), labelled here with $\varphi(.)$. Such a description of the likelihood helps to deal with correlated heteroscedastic model errors by transferring them into the space where they can be assumed as homoscedastic. Thus, the joint distribution of the runoff $\boldsymbol{Y}$ conditional on all parameters $\boldsymbol{\theta}$, σ, τ can be described with the following multivariate Gaussian distribution:

$$p(\boldsymbol{y}^o|\boldsymbol{\theta},\sigma,\tau,\boldsymbol{x}) = \frac{(2\Pi)^{-n/2}}{\sqrt{|\boldsymbol{\Sigma}(\sigma,\tau)|}} \cdot \exp\left(-\frac{1}{2}[\tilde{Y}_i(\boldsymbol{\theta},x_i)]' \cdot \boldsymbol{\Sigma}^{-1}(\sigma,\tau) \cdot [\tilde{Y}_i(\boldsymbol{\theta},x_i)]\right) \quad (3)$$

with a covariance matrix $\boldsymbol{\Sigma}$ expressed as

$$\boldsymbol{\Sigma}(\sigma,\tau) = \sigma^2 \cdot (1 - \exp(2 \cdot \omega_i(\tau)) \quad (4)$$

where

$$\omega_i(\tau) = |t_i - t_{i-1}|/\tau \quad (5)$$

$$\tilde{Y}_i(\boldsymbol{\theta},x_i) = \left[\varphi(y_i^o) - \varphi\left(\hat{y}_i(\boldsymbol{\theta},x_i)\right)\right] \quad (6)$$

$\tilde{Y}_i(\boldsymbol{\theta},x_i)$ are residuals of the model at time step t_i in the transformed space. x_i are input variables, $\varphi(y_i^o)$ and $\varphi\left(\hat{y}_i(\boldsymbol{\theta},x_i)\right)$, stand for observations and the deterministic output of the HBV model after applying the Box-Cox transformation, $\varphi(.)$.

Computation of this likelihood for long observations of $\boldsymbol{y}^o$ (i.e. having numerous data points) is computationally expensive and requires inverting large matrices. To avoid this problem, we used a conditional likelihood function instead which is computed at each time step and thus avoids inverting the matrix. Such conditional likelihood takes the form of:

$$p(\boldsymbol{y}^o|\boldsymbol{\theta},\sigma,\tau,\boldsymbol{x}) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2\sigma^2}\tilde{Y}_0^2(\boldsymbol{\theta},x_0)\right) \cdot \prod_{i=1}^{n}\left[\left(\sqrt{2\pi}\cdot\sigma\cdot\sqrt{1-\exp\left(-2\omega_i(\tau)\right)}\right)^{-1} \cdot \exp\left(-\frac{\left[\tilde{Y}_i(\boldsymbol{\theta},x_i)-\tilde{Y}_{i-1}(\boldsymbol{\theta},x_{i-1})\cdot\exp\left(-\omega_i(\tau)\right)\right]^2}{2\sigma^2\left[1-\exp\left(-2\omega_i(\tau)\right)\right]}\right) \cdot \left|\frac{d\varphi}{dy}(y_i^o)\right|\right] \quad (7)$$

In Equation (7), *n* and *i* represent the length of and the subscript over the calibration period. The uppercase stands for a random variable and lowercase for a constant variable. Use of such a conditional likelihood function is a convenient approach and has been previously applied in literature (e.g. Yang et al., 2007, Sikorska et al. 2012).

3.5 Descriptors of the model performance

3.5.1 Efficiency criteria

The overall model performance was assessed with a commonly used Nash-Sutcliffe efficiency, R_{NS} (-) (Krause *et al.* 2005) supported by a Kling-Gupta efficiency, R_{KGE} (-) (Gupta *et al.* 2009):

$$R_{\mathrm{NS}} = 1 - \left(\sum_{i=1}^{n}(\hat{y}_i - y_i^o)^2 / \sum_{i=1}^{n}(y_i^o - \bar{y}^o)\right) \quad (8)$$

where *n* indicates the number of observation points, *i* is the *i*th observation point.

$$R_{\mathrm{KGE}} = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2} \quad (9)$$

where *r* is the linear correlation coefficient between $\boldsymbol{y}^o$ and $\hat{\boldsymbol{y}}$, *a* is a measure of the relative variability in $\boldsymbol{y}^o$ and $\hat{\boldsymbol{y}}$, and *b* is the bias that represents the ratio between the mean simulated ($\bar{\hat{\boldsymbol{y}}}$) and mean observed ($\bar{\boldsymbol{y}}^o$) variables.

The R_{NS} and R_{KGE} were assessed in calibration and validation always only for the best model simulation, i.e. corresponding to the parameter set receiving the highest value for the likelihood function during inference.

3.5.2 Metrics of uncertainty bands

Model simulations in the validation period were assessed with two quantitative metrics: the uncertainty coverage of data points, R_{DC} (%) and the average uncertainty band spread, R_{ABS} (mm h^{-1}).

$$R_{\mathrm{DC}} = 100 \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} i \quad \text{where: } \begin{cases} i = 1, & if\ y_i^o \in \left[\hat{y}_i^L; \hat{y}_i^U\right] \\ i = 0, & \text{else} \end{cases} \quad (10)$$

$$R_{\mathrm{ABS}} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\hat{y}_i^U - \hat{y}_i^L \right) \quad (11)$$

where $\hat{y}_i^U$ and $\hat{y}_i^L$ are the upper and lower limits of the uncertainty bands at time point *i*, here taken as 95% and 5%.

3.5.3 Benchmark: Nash-Sutcliffe optimization

To compare the model performance at different window lengths of precipitation aggregation and disaggregation, we used a benchmark model simulation. The benchmark was computed by a traditional optimization approach, i.e. by minimizing the least squares error, using the Nash-Sutcliffe efficiency as a single model performance criterion. Note that this optimization approach corresponds to assuming homoscedastic (normally distributed) model errors, as discussed in Section 3.3.2. This unrepresented model error is expected to translate to model parameters.

3.6 Prior settings for the Bayesian inference

As dependences between model parameters cannot be known in advance, we formulated the joint prior on all model and error parameters as an independent univariate normal distribution. For each parameter of the HBV model we used a uniform distribution restricted to possible values. The prior formulation of the error model is difficult because it is likely to depend on many factors (Renard *et al.* 2011). Thus, for its parameters (σ and τ), we used a truncated normal and a log-normal distribution which are bounded towards zero (Table 2). Such an error formulation aims first at obtaining a maximum possible performance with the deterministic runoff model, while the error model is next adjusted to capture the remaining variability in observation data. In this way, the overestimation of the error model should be avoided. Thanks to such a general prior formulation, it is directly transferable to other mesoscale catchments.

These prior parameters are updated during each calibration scheme to posterior by executing the conditional likelihood (Section 3.4.2.). Practically, this is done by directly sampling from the posterior using a Markov Chain Monte Carlo (MCMC) algorithm which is computationally challenging and requires efficient strategies (Thyer et *al.* 2009). Possible strategies for implementing MCMC algorithms are reported in literature elsewhere (Kuczera *et al.* 2007, Renard *et al.* 2009, Sikorska *et al.* 2012). For each calibration scheme, observation periods from 2005 to 2008 were used to calibrate the stochastic runoff model with a six-month pre-processing warming-up period, while years 2009–2010 were used to validate the model.

4 Results

4.1 Model identification at different temporal averaging window lengths

The stochastic runoff model was identified slightly differently at different averaging window lengths (*α)* which resulted in parameter posteriors slightly shifted as compared to each other (not shown). All posteriors were also much narrower than generic priors and for most of the HBV parameters they overlapped. On the contrary, parameters of the error model varied more, especially the standard deviation of the model error (σ) which was also identified from data at higher values than assumed a priori. Due to interdependencies between all inferred parameters, analysis of parameter groups instead of individual parameters seems more appropriate. Diagnosing four different model components, i.e. (a) response and routing, (b) snow, (c) soil and (d) error, we found that the relative contribution of the error component slightly decreased with enlarging α, while the contribution of the response and routing, and snow slightly increased (see Fig. 3). For the soil component, changes in its contribution along α have been observed but without any strong patterns.

Despite the fact that most posterior parameters were identified similarly, model simulations varied depending on the chosen α for precipitation averaging and source of precipitation data (SNP or RBP). The resulting model efficiencies assessed with the Nash-Sutcliffe (R_{NS}) and Kling-Gupta (R_{KGE}) measures are plotted for calibration period in Figure 4. As can be seen, averaging over periods of up to a few hours usually did not have any significant effect on the model performance (except catchment C9). After that period, a gradual decrease in model performance could be observed with lengthening α. Comparing two different datasets, a better model performance was observed with radar-based precipitation (RBP) than with station network precipitation (SNP) for all of the catchments. Moreover, when using RBP data, longer window lengths were suitable for providing a good model performance than when using SNP data. In addition, although similar patterns for enlarging α were observed for both measures i.e. R_{NS} and R_{KGE}, generally higher values were achieved for R_{KGE} than R_{NS}.

When comparing results of the Bayesian inference with the benchmark approach (standard calibration with a Nash-Sutcliffe criterion used as the objective function), similar model efficiencies were obtained in most catchments for both datasets.

4.2 Effect of precipitation temporal averaging on model predictive performance

Results from the independent period revealed similar patterns to those observed in calibration, meaning that lengthening α decreased the model performance (Fig. 5). Also, the dataset based on RBP gave a better model performance at the same averaging window that

it was obtained with the SNP dataset and this effect became more pronounced in validation and was more visible for R_{NS} than for R_{KGE}. Moreover, for some catchments only a high resolution data led to an acceptable model performance (C2 and C9).

Diagnosing the behaviour of uncertainty bands, we generally found that, on average, extending the length of α did not have any significant effect on the average band spread (R_{ABS}) of simulations with RBP and SNP data (Fig. 6). However, with averaging longer than 10 h, it did gradually decrease the percentage of data points lying within the uncertainty bands (R_{DC}), and this effect was better visible for simulations using SNP data. In this context, uncertainty bands for simulations using RBP data were usually slightly wider than those of SNP but they were more reliable because they covered more data-points.

4.3 Diagnosis of the model performance in the context of the catchment grouping

When looking at catchments grouped by size, two patterns can be noticed. First, the model performance decreased from large to small catchments and from short to longer averaging windows (α). Thus, the length of the averaging window played a larger role in small (A < 60 km^2) than in medium (60 ≤ A <200 km^2) and large (A ≥ 200 km^2) catchments (Fig. 7).

Second, precipitation averaging based on RBP resulted in a higher model performance than the one that used SNP. This difference was especially pronounced in small catchments, where SNP data did not result in a good model performance assessed by the Nash-Sutcliffe efficiency (with R_{NS} ≤ 0.5). In medium catchments, longer averaging windows were allowed when using RBP than SNP data. Thus, using the averaging approach together with the RBP datasets allowed the window length α to be increased up to 10–12 h in small catchments, to 15–21 h in medium catchments, while preserving an acceptable model predictive performance (R_{NS} ≥ 0.5). When using the averaging based on SNP data, smaller averaging windows were required: maximum 1 hour for small catchments (observed data) and 10–12 h for medium catchments. However, averaging precipitation in small catchments generally significantly decreased model performance resulting in R_{NS} ≤ 0.5. Interestingly, for large size catchments, both datasets resulted in a similarly good performance with a slight advantage of the SNP dataset, and the permitted window length was up to 24 h. Confronting these findings with the Kling-Gupta efficiency (Fig. 8) demonstrated that this metric was less sensitive to changes in α and usually longer averaging windows were suitable for an acceptable model performance (R_{KGE} ≥ 0.5).

Diagnosis of the model error (σ) revealed that higher values of the model error were estimated in smaller and lower values in medium and large catchments. However, σ was only insignificantly varied over the averaging window resulting in a slightly higher value for small

averaging windows (Fig. 9). Model error was also most often higher in averaging approaches using SNP than those based on RBP data especially for small catchments. In contrast, the characteristic correlation length (r) was less sensitive to lengthening the averaging window (not shown) and similar values were obtained for all window lengths. For the benchmark no model error could be estimated (because the model error is not included in the parameter inference).

5 Discussion

5.1 Precipitation temporal resolution versus catchment size

Results from our inter-catchment comparison showed that lengthening the averaging window of precipitation data (α) gradually decreased the model performance and this effect was data-source and catchment-size dependent. While small catchments (<60 km^2) were very sensitive to the length of α, medium catchments (60 ≤ A < 200 km^2) were less sensitive and large catchments (≥200 km^2) were least affected by changes in α. It is clear that the time scale of runoff processes is linked to the catchment size and in larger catchments these processes may occur over a few days, while in smaller catchments over a few hours. Hence, as it could be expected, daily precipitation data does not provide enough information to capture and represent runoff dynamics in catchments of a response time smaller than one day.

Our results also demonstrated that timely averaging of precipitation data combined with a high model resolution had a lower impact on the parameter identification and a higher effect on the model performance. This effect is likely to be caused by the applied error model which compensates for uncertainties due to precipitation averaging and in this way allows the effect of data averaging on model parameters (see further below) to be minimized, as it was also suggested by Kavetski *et al.* (2011).

5.2 Station vs *radar-based precipitation: model performance*

Comparison of radar-based (RBP) with station network precipitation (SNP) highlighted the advantage of RBP data, an observation in agreement with our previous work which however focused on only one catchment and used a different error model (Sikorska and Seibert 2016). The benefit of using the RBP data was mostly pronounced in small and medium catchments (i.e. with area ≤ 200 km^2). One reason for this is a small set of stations usually available for small catchments, while this set increases with increasing the catchment size because stations from a larger area can be used. In contrast, in small catchments, in the absence of stations located nearby, remotely located stations often have to be used.

Another point is linked to the catchment response to precipitation events. The specific response will depend largely on the precipitation totals, their temporal intensities (here averaging window length), and the location of the main mass of the rainfall field and its trajectory through the catchment (Theyer *et al.* 2009). As these effects are smoothed up in larger catchments, the SNP datasets are capable of providing more realistic representation of the precipitation event than it is possible in smaller catchments, where local extrema may play a significant role. Given a small set of stations in small catchments, it is likely that these extrema will be missed by the sparse network.

In this aspect, RBP, which incorporates observations from weather radars, provides a more reliable source of average precipitation totals, and is able to better capture spatial precipitation information across the entire catchment area. Two issues must be however highlighted here. First, that the RBP uses information from both weather radars and ground stations, which are used to correct radar estimates into precipitation totals and thus cannot be applied without ground stations. Second, SNP data is used in this study in combination with a Thiessen polygon method for spatially averaging point information from stations. Thus, the way of averaging such information may have an effect on the observed results and this is discussed further in Section 5.6. It is also clear that the advantage of RBP over SNP became less visible in larger catchments due to the reasons mentioned above.

5.3 Station vs *radar-based precipitation: uncertainties*

A better model performance obtained with RBP data was also confirmed by the diagnosis of the inferred model errors which were on average higher: (a) for calibration schemes using SNP; (b) in small catchments; and (c) for finer precipitation resolutions (shorter averaging windows). Because in each averaging scheme the same runoff data and the same runoff model were used, the differences in estimated model errors can be explained due to precipitation uncertainty.

Thus, precipitation uncertainty was generally higher for finer temporal resolutions in all catchment groups. This is an interesting finding which suggests that averaging precipitation totals may indeed smooth precipitation errors due to measurement, spatial and temporal representation of the catchment mean areal precipitation, or other random and systematic errors. Similar patterns in decreasing precipitation uncertainty due to lengthening temporal averaging intervals were recently reported by Muthusamy *et al.* (2017).

Another discussion point relates to the use of the benchmark. As our results showed, a traditional calibration approach may result in a similarly good model performance (although violating assumptions on normally distributed errors). Thus introducing model error does not

necessarily lead to a better model performance assessed by standard performance metrics (i.e. Nash-Sutcliffe or Kling-Gupta). Yet, it enables model errors to be represented and resulting uncertainty to be quantified in a structural way, leading to more reliable estimates.

5.4 Precipitation temporal resolution as a control of the runoff model performance

The above findings have further implications for hydrological applications. As our results showed, the temporal resolution of averaged data appears to be a control metric for the model performance as it allows an optimal resolution for hydrological modelling to be selected. This is an important finding for catchments with two types of the data availability, i.e. (a) when hourly datasets are available but could be aggregated to minimize the amount of data points for advancing model simulations, or (b) when only sparse data (e.g. daily) are available which could be disaggregated to finer resolutions to help identify model parameters at high temporal scales. In both cases, our work provides a framework for selecting such an optimal temporal resolution (averaging window length) of precipitation data in a structural way.

From our catchment grouping analysis, we were able to identify thresholds suitable for applying precipitation averaging. For datasets relying on RBP these thresholds were: 10-12 h for small catchments, and up to 21 h for medium size catchments, and 24 h for large catchments. These thresholds were shorter when using the SNP data (apart from large catchments for which similar window lengths were required).

Interestingly, this finding was metric-dependent and using the Kling-Gupta efficiency instead of the Nash-Sutcliffe efficiency resulted in allowing the use of much longer averaging windows. This indicates that the choice of the appropriate precipitation aggregation length should be influenced by the focus of simulation, i.e. by choosing a performance measure adequately to the hydrograph characteristic which should be simulated.

5.5 Methodological aspects

It has to be stressed however that some aspects observed in this study are linked with the models applied in this study (HBV and error model). For instance, the effect of lengthening the window length on parameters of the HBV model was poorly visible. This effect could be however more pronounced for other hydrological models that describe runoff processes in more detail, e.g. represent fast flow components such as overland flow. As this process is not represented within HBV, such effects cannot be detected. Also, the error model used here represents all the remaining uncertainty terms not described with parameters of the hydrological model (HBV). As only precipitation inputs are altered between different

calibration schemes, we assume that the difference in the observed error model behaviour is due to increasing/decreasing precipitation uncertainty. This assumption is valid under the assumption of the linearity of uncertainty components, and particularly of the proportionality of the total uncertainty to the increasing/decreasing precipitation uncertainty. Such an assumption is justified when only relative contributions are of interest, as in this work, and is common in the hydrological community (e.g. Kuczera *et al.* 2010, Renard *et al.* 2011, Sikorska *et al.* 2012). Finally, due to a small set of studied catchments and to minimize effects of local conditions, the analysis should ideally be re-evaluated at a larger sample of catchments. Due to the computational demands, such an analysis is (currently) not possible on a standard PC and would require cloud or cluster computations.

5.6 Use of Thiessen polygons for computing mean areal precipitation from SNP data

It is worth noting that our results regarding SNP data are purely related to the method of obtaining mean areal precipitation inputs for the hydrological model. In our case this was a commonly applied Thiessen polygons' method. Although this method was criticized for being too simple and more advanced methods of precipitation averaging were proposed, due to its simplicity and flexibility in applications at different catchment conditions, it is still one of the most commonly applied method in hydrological modelling to average point precipitation records available from ground stations. To evaluate the sensitivity of our results to the method for precipitation areal averaging, we performed a simple simulation study, where precipitation inputs obtained from SNP data were corrupted with a random error sampled from a normal distribution with a mean equal 0 and a standard deviation of 0.1. Results of this experiment performed for three study catchments from different catchment group and a poor model performance are presented in Figure 10, which shows two points. First, although introducing a random error may lead to different (partly improved) model efficiency metrics (assessed by R_{NS} and R_{KGE}), the previously observed patterns for different aggregation schemes remained similar. Second, the RBP dataset still provided the best model simulations. Based on these findings, we argue that a low density of raingauges and their location down in valleys rather than the method of obtaining the mean areal catchment precipitation are major sources contributing to a poorer model performance with the SNP than with the RBP dataset. This is especially of concern for mountainous catchments with a difficult terrain that disturbs the accuracy of precipitation totals measured with a point gauges and their and representativeness at the catchment scale.

6 Conclusions

The approach used in this study provides a structural way for quantifying the effects of a precipitation temporal averaging approach on the model performance and associated uncertainty. Our results demonstrated that a high hourly temporal resolution may not always be required and optimally averaged precipitation inputs may be capable of providing acceptable performance simulations. The averaging window length is catchment size and data source dependent and in our case we found that temporal resolutions of up to 1–12 h were suitable for small catchments (<60 km^2), up to 12–21 h for medium size catchments (<200 km^2), and up to 24 h for large (≥200 km^2) catchments. The lower values refer to SNP while larger thresholds were found for the RBP dataset. Our results also indicated that RBP has a higher value for deriving precipitation data in catchments with area smaller than 200 km^2, while this effect vanishes in larger catchments. These quantitative results from our study are useful for selecting an optimal temporal resolution of precipitation data and the precipitation data source to support parameter identification in mesoscale pre-alpine catchments and can be used to decide on aggregating hourly datasets or disaggregating sparse data.

Acknowledgements

The Swiss Federal Office for the Environment (FOEN) and MeteoSwiss are acknowledged for providing runoff and meteorological data. The authors are grateful to Tracy Ewen (University of Zurich) for proofreading the manuscript. The authors thank the associate editor Giuliano Di Baldassarre, Korbinian Breinl, and one anonymous reviewer for their useful comments which helped us improve this manuscript. The ScienceCloud provided by S3IT at the University of Zurich enabled the computational-intensive simulations to be run on virtual machines.

Funding

This work was supported by the Polish Ministry of Science and Higher Education within the program "Mobility Plus" [Grant no.1097/MOB/2013/0].

References

Aronica, G., Freni, G. and Oliveri, E., 2005. Uncertainty analysis of the influence of rainfall time resolution in the modelling of urban drainage systems. *Hydrological Processes,* 19, 1055-1071.

Bastola, S. and Murphy, C., 2013. Sensitivity of the performance of a conceptual rainfall-runoff model to the temporal sampling of calibration data. *Hydrology Research*, 44, 3, 484-494, doi:10.2166/nh.2012.061

Bergström, S., 1992. The HBV model - its structure and applications, *SMHI Hydrology*, RH No.4, Norrkping, 35 pp.

Blöschl, G., Reszler, C., and Komma, J., 2007. A spatially distributed flash flood forecasting model, *Environmental Modelling & Software*, 23, 464–478.

Booij, M. J., 2002. Modelling the effects of spatial and temporal resolution of rainfall and basin model on extreme river discharge, *Hydrological Sciences Journal,* 47:2, 307-320, doi: 10.1080/02626660209492932.

Breinl, K., 2016. Driving a lumped hydrological model with precipitation output from weather generators of different complexity, *Hydrological Sciences Journal*, 61, 1395-1414, 10.1080/02626667.2015.1036755.

Gelman, A., Carlin, J.B., Hal S.S., Dunson, D.B., Vehtari, A. and Rubin, D.B., 2013. *Bayesian Data Analysis*, pp. 675, Chapman & Hall, London.

Germann, U., Galli, G., Boscacci, M., Bolliger, M., 2006. Radar precipitation measurement in a mountainous region, *Quarterly Journal of the Royal Meteorological Society,* 132, 1669–1692 http://dx.doi.org/10.1256/qj.05.190.

Griessinger, N., Seibert, J., Magnusson, J., and Jonas, T., 2016. Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments, *Hydrology and Earth System Sciences*, 20, 3895-3905, doi:10.5194/hess-20-3895-2016.

Gupta, H. V., Kling, H., Yilmaz, K. K., Martinez, G. F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology*, 377, 1–2, 80-91, http://doi.org/10.1016/j.jhydrol.2009.08.003.

Hearman, A.J. and Hinz, C., 2007. Sensitivity of point scale surface runoff predictions to rainfall resolution, *Hydrology and Earth System Sciences*, 11, 965-982, doi:10.5194/hess-11-965-2007.

Joss, J. and R. Lee, R., 1995. The application of radar radar-gauge comparisons to operational precipitation profile corrections, *Journal of Applied Meteorology*, 34, 2612-2630.

Jost, G., Moore, R. D., Menounos, B., and Wheate, R., 2012. Quantifying the contribution of glacier runoff to streamflow in the upper Columbia River Basin, Canada, *Hydrology and Earth System Sciences*, 16, 849-860, doi:10.5194/hess-16-849-2012.

Kavetski, D., Fenicia, F. and Clark, M.P., 2011. Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment, *Water Resources Research*, 47, W05501, doi:10.1029/2010WR009525.

Koutsoyiannis, D., and Onof, C., 2001. Rainfall disaggregation using adjusting procedures on a Poisson cluster model, *Journal of Hydrology*, 14, 109-122, http://dx.doi.org/10.1016/S0022-1694(01)00363-8.

Kossieris, P., Makropoulos, C., Onof, C. and Koutsoyiannis, D., 2016. A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett-Lewis based model with adjusting procedures, *Journal of Hydrology*, http://dx.doi.org/10.1016/j.jhydrol.2016.07.015.

Krause, P., Boyle, D.P., and B'ase, F., 2005. Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences*, 5, 89-97, doi:10.5194/adgeo-5-89-2005.

Kuczera, G., Renard, B., Thyer, M. and Kavetski, D., 2010. There are no hydrological monsters, just models and observations with large uncertainties! *Hydrological Sciences Journal,* 55, 980-991. http://dx.doi.org/10.1080/02626667.2010.504677.

Littlewood, I.G., and Croke, B.F.W., 2008. Data time-step dependency of conceptual rainfall-streamflow model parameters: an empirical study with implications for regionalisation, *Hydrological Sciences Journal*, 53:4, 685-695, DOI: 10.1623/hysj.53.4.685

Muthusamy, M., Schellart, A., Tait, S., and Heuvelink, B.M. G. 2017. Geostatistical upscaling of rain gauge data to support uncertainty analysis of lumped urban hydrological models, *Hydrology and Earth System Sciences,* 21, 1077-1091, https://doi.org/10.5194/hess-21-1077-2017.

Nans, A., Rössler, O., Köplin, N., Huss, M., Weingartner, R., and Seibert, J. 2014. Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. *Water Resources Research*, 50(10):7541-7562. doi:10.1002/2014WR015549.

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R.D., Reinoso-Rondinel, R., Bruni, G., et al., 2015. Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation, *Journal of Hydrology*, 531, 2, 389-407, doi: http://dx.doi.org/10.1016/j.jhydrol.2015.05.035.

Patil, S.D. and Stieglitz, M., 2015. Comparing spatial and temporal transferability of hydrological model parameters, *Journal of Hydrology*, 525, 409-417, http://dx.doi.org/10.1016/j.jhydrol.2015.04.003.

Pui A., Sharma, A., Mehrotra, R., Sivakumar, B. and Jeremiah, E., 2012. A comparison of alternatives for daily to sub-daily rainfall disaggregation, *Journal of Hydrology*, 470-471, 138-157, doi:http://dx.doi.org/10.1016/j.jhydrol.2012.08.041.

Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and Franks, S.W., 2011. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resources Research*, 47, W11516, doi:10.1029/2011WR010643.

Reynolds, J.E., Halldin, S., Xu, C.Y., Seibert, J., and Kauffeldt, A., 2015. Sub-daily runoff simulations with parameters inferred at the daily time scale, *Hydrology and Earth System Sciences Discuss.,* 12, 7437-7467, doi:10.5194/hessd-12-7437-2015, 2015.

Seibert, J., 1997. Estimation of parameter uncertainty in the HBV model, *Nordic Hydrology*, 28 (4/5), pp. 247-262.

Sikorska, A.E., Scheidegger, A., Banasik, K. and Rieckermann, J., 2012. Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models. *Hydrology and Earth System Sciences,* 16, 1221-1236, doi:10.5194/hess-16-1221-2012.

Sikorska, A.E, Montanari, A. and Koutsoyiannis, D., 2015a. Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. *Journal of Hydrologic Engineering*, 20(A4014009); DOI: 10.1061/(ASCE)HE.1943-5584.0000926.

Sikorska, A.E., Viviroli, D., and Seibert, J., 2015b. Flood-type classification in mountainous catchments using crisp and fuzzy decision trees, *Water Resources Research*, 51, doi:10.1002/2015WR017326.

Sikorska, A.E., Del Giudice, D., Banasik, K. and J. Rieckermann, 2015c. The value of streamflow data in improving TSS predictions -- Bayesian multi-objective calibration, Journal of Hydrology, 530, 241-254, doi: http://dx.doi.org/10.1016/j.jhydrol.2015.09.051.

Sikorska, A.E. and Seibert, J., 2016. Value of different precipitation data for flood prediction in an alpine catchment: A Bayesian approach, *Journal of Hydrology*, doi:10.1016/j.jhydrol.2016.06.031.

Staudinger, M., Weiler, M., and Seibert, J., 2015. Quantifying sensitivity to droughts - an experimental modeling approach, *Hydrology and Earth System Sciences,* 19, 1371-1384, doi:10.5194/hess-19-1371-2015.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological

modeling: A case study using Bayesian total error analysis, *Water Resources Research,* 45, W00B14, doi:10.1029/2008WR006825.

Wang, Y., He, B. and Takase, K., 2009. Effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge, *Hydrological Sciences Journal*, 54:5, 886-898, DOI: 10.1623/hysj.54.5.886.

Wetterhall, F., He, Y., Cloke, H. and Pappenberger, F., 2011. Effects of temporal resolution of input precipitation on the performance of hydrological forecasting, *Advances in Geosciences*, 29, 21-25, doi:10.5194/adgeo-29-21-2011.

Table 1. Properties of the study catchments sorted by the increasing catchment size.

ID	River, Gauging site	Area (km^2)	Mean elevation (m a.s.l.)	Regime type[b]	Precipitation		Runoff[d] (mm $year^{-1}$)
					SNP[c] (mm $year^{-1}$)	RBP[c] (mm $year^{-1}$)	
Small[a]							
C1	Glatt, Herisan	16.2	840	P	1827	1447	1078
C2	Sionge-Vuippens, Chateau	45.3	862	Np	1061	1232	714
C3	Alp, Einiedeln	46.4	1155	Np	2196	1890	1539
C4	Goldach-Goldach, Bleiche	49.8	833	P	1830	1324	913
C5	Minster, Euthal	59.0	1318	N	2485	2055	1597
Medium							
C6	Murg, Wängi	78.9	650	P	1578	1246	675
C7	Emme, Eggiwil	124.0	1249	Np	1488	1664	1032
C8	Suze, Sonceboz	150.0	1076	Np	768	1292	859
C9	Birse, Moutier	183.0	907	Np	776	1372	536
C10	Ilfis, Langnau	188.0	1002	Np	1005	1668	880
Large							
C11	Plessur, Chur	263.0	1917	N	1138	1024	843
C12	Emme, Emmenmatt	443.0	1004	Np	1227	1575	849
C13	Emme, Wiler	939.0	841	P	1079	1378	632

[a] Catchment grouping according to area, A: Small: $A < 60$ km^2; Medium: $60 \leq A < 200$ km^2; and Large: $A \geq 200$ km^2.

[b]River regime types: p: pluvial; n: nival; and np: nivo-pluvial; based on Weingartner and Aschwanden (1992).

[c,d] Mean annual areal catchment precipitation and mean annual runoff over the recorded period; SNP: station network; RBP: radar-based precipitation.

Table 2 Initial probability distributions for the parameter inference.

Parameter[a]	Min.	Max.	Description
HBV snow routine:			
TT (°C)	-5	2.5	Threshold temperature
CFMAX (mm h^{-1} °C^{-1})	0.01	10	Degree-hour factor
SFCF (-)	0.4	1.6	Snowfall correction factor
CFR (-)	0	0.1	Refreezing correction factor
CWH (-)	0	0.2	Water holding capacity
HBV soil moisture routine:			
CET (°C^{-1})	0	0.5	Correction factor for potential evaporation
FC (mm)	50	550	Maximum moisture storage in soil box
LP (-)	0.1	1	Threshold for reduction of evaporation
BETA (-)	0.1	10	Shape coefficient
HBV response routine:			
PERC (mm h^{-1})	0	4	Percolation parameter
UZL (mm)	0	100	Groundwater runoff threshold parameter
K0 (h^{-1})	0.00001	0.4	Recession coefficient[b]
K1 (h^{-1})	0.000001	0.2	Recession coefficient[b]
K2 (h^{-1})	0.0000001	0.1	Recession coefficient[b]
HBV routing routine:			
MAXBAS (h)	1	100	Length of triangular weighting function
Error term:	Mean	SD	
σ (φ(mm h^{-1}))	0	$\varphi(0.001\ y^{o}_{max})$	Asymptotic standard deviation
τ (h)	24	12	Characteristic correlation length

[a]Applied distributions: Uniform for HBV parameters; truncated normal for σ; and lognormal for τ, where φ is the Box-Cox transformation function, y^{o}_{max} is the maximum runoff over the analysed period.

[b]If lower than the threshold, a respective upper box is activated.

Figure captions

Figure 1. Location of the 13 study catchments. Colours indicate the catchment size, the mesh presents Thiessen polygons estimated for hourly precipitation stations; dots: locations of weather radars.

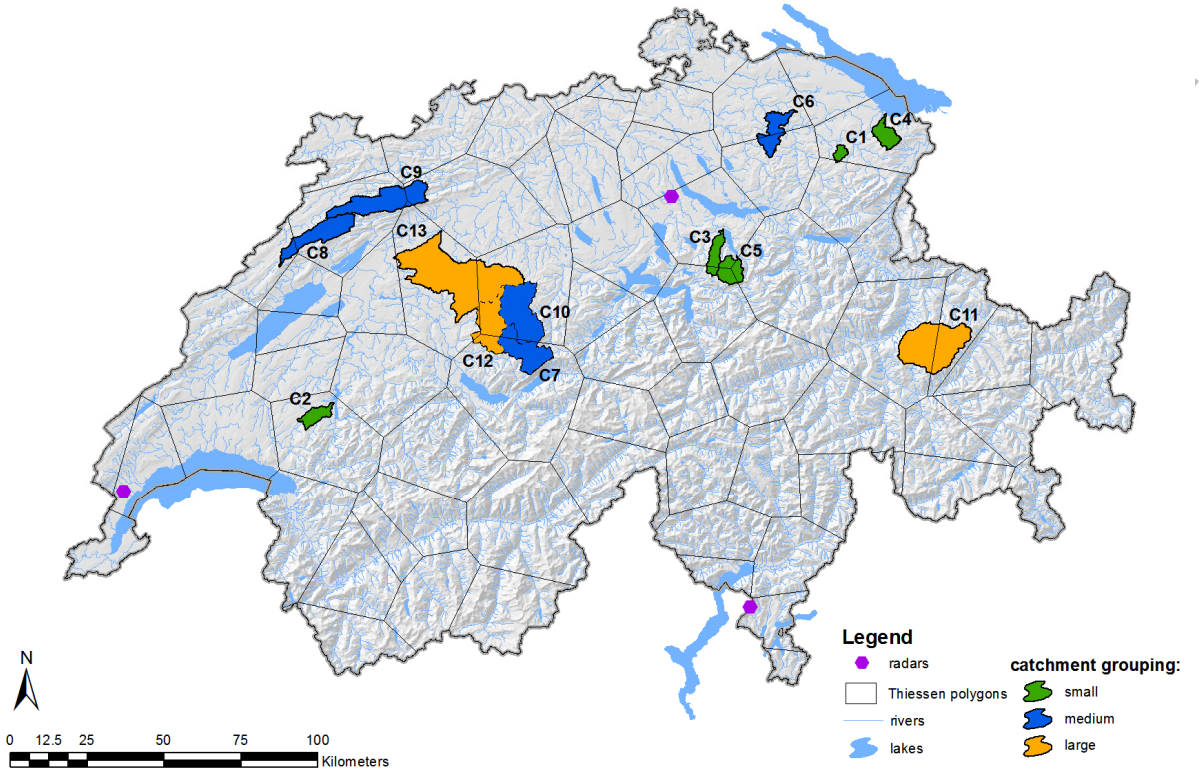

Figure 2. Schematic overview of the data temporal averaging approach.

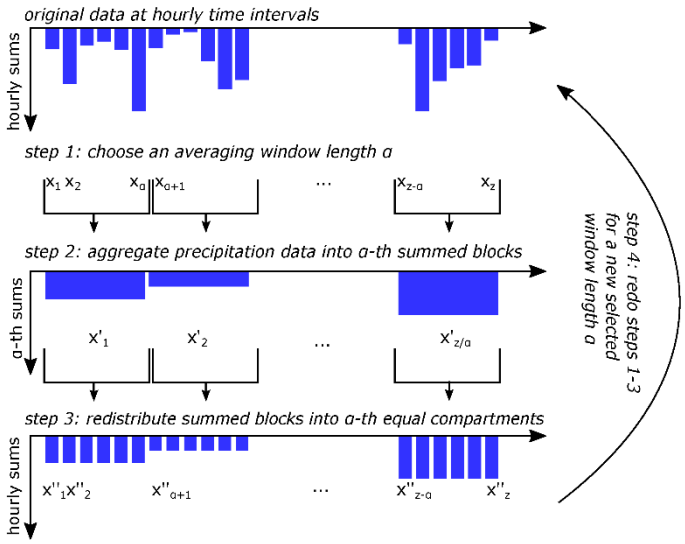

Figure 3. Posterior parameter analysis: relative contribution of model parameters, grouped into four main components, i.e. response and routing, soil, snow and error, to the averaging

window length, α (h), for all study catchment and the SNP (top) and RBP (bottom) datasets. A slight decrease in the error contribution and an increase in the response and routing contribution with an increasing α can be observed.

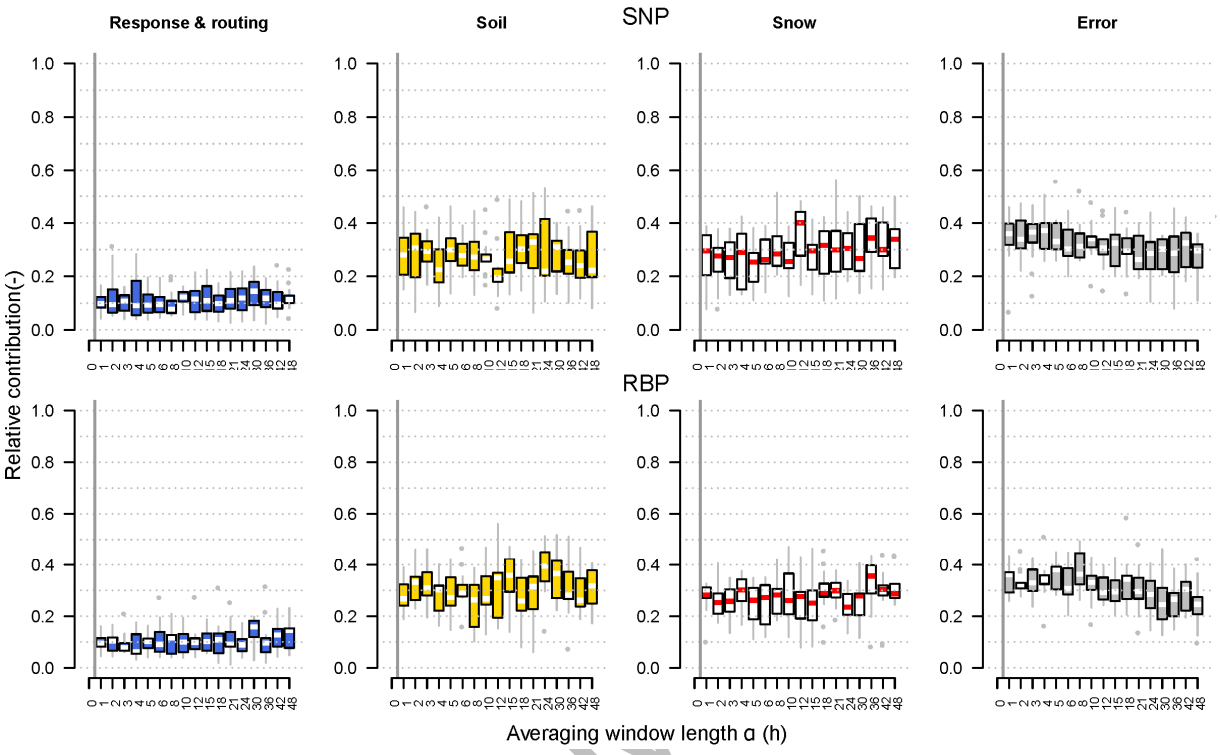

Figure 4. Nash-Sutcliffe efficiency, R_{NS} (top) and Kling-Gupta efficiency, R_{KGE} (bottom) in calibration for 13 study catchments (C1–C13) and for different averaging window lengths, α (1–48 h) and two precipitation datasets (SNP and RBP). The α labelled "B" corresponds to the benchmark.

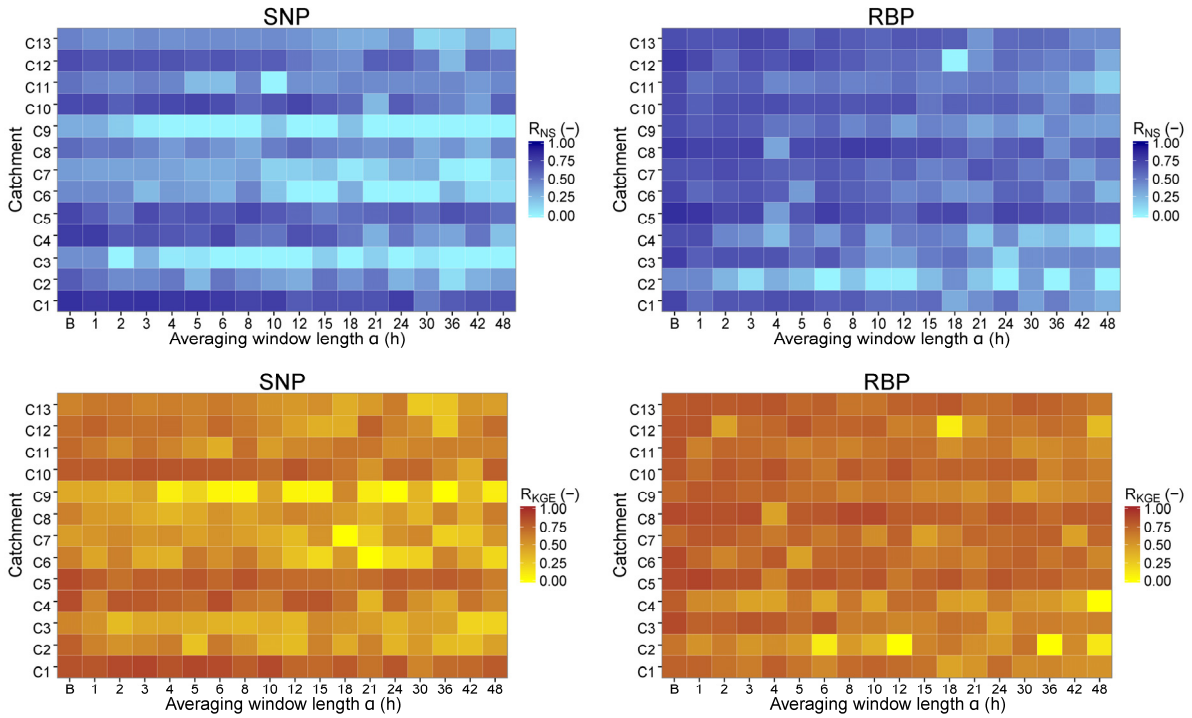

Figure 5 Nash-Sutcliffe efficiency, R_{NS} (top) and Kling-Gupta efficiency, R_{KGE} (bottom) in validation for 13 study catchments (C1–C13) and for different averaging window lengths α (1–48 h) and two precipitation datasets (SNP and RBP). The α labelled "B" corresponds to the benchmark.

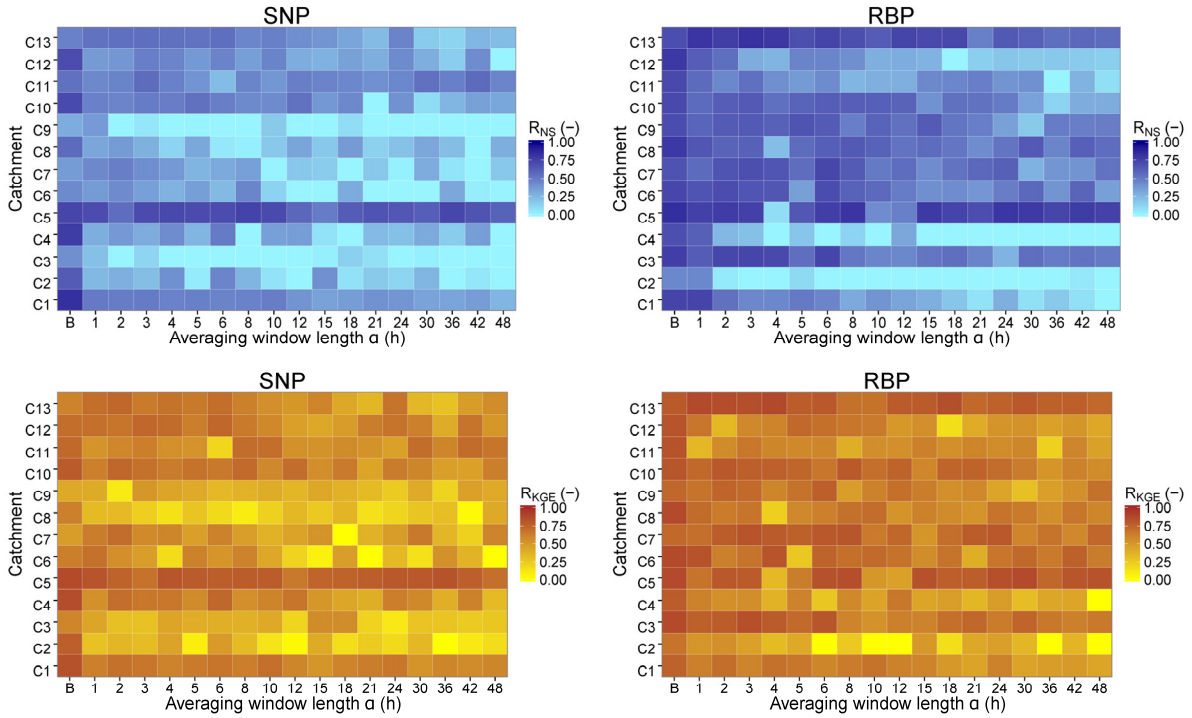

Figure 6. Data coverage, R_{DC} (top) and average band spread, R_{ABS} (bottom) of uncertainty bands *vs* the averaging window length α (1–48 h) in the validation period for 13 study catchments and two precipitation datasets (SNP and RBP). The α labelled "B" corresponds

to the benchmark. Note that for the benchmark no estimation of uncertainty bands can be given.

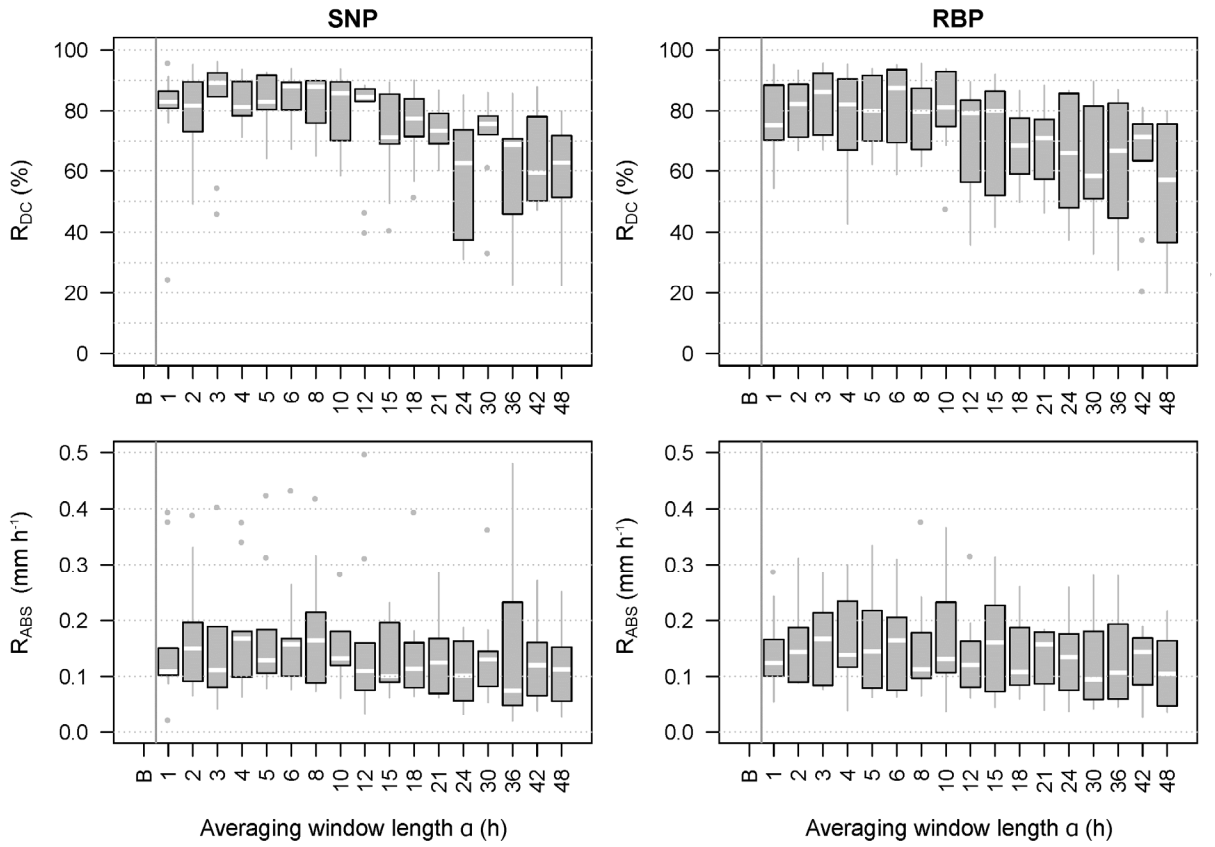

Figure 7. Averaged effect of the data temporal averaging (α) on model identification assessed by the Nash-Sutcliffe efficiency with two precipitation datasets (RBP and SNP) grouped according to catchment area into small (A < 60 km^2); medium (60 ≤ A < 200 km^2); and large (A ≥ 200 km^2). The α labelled "B" corresponds to the benchmark.

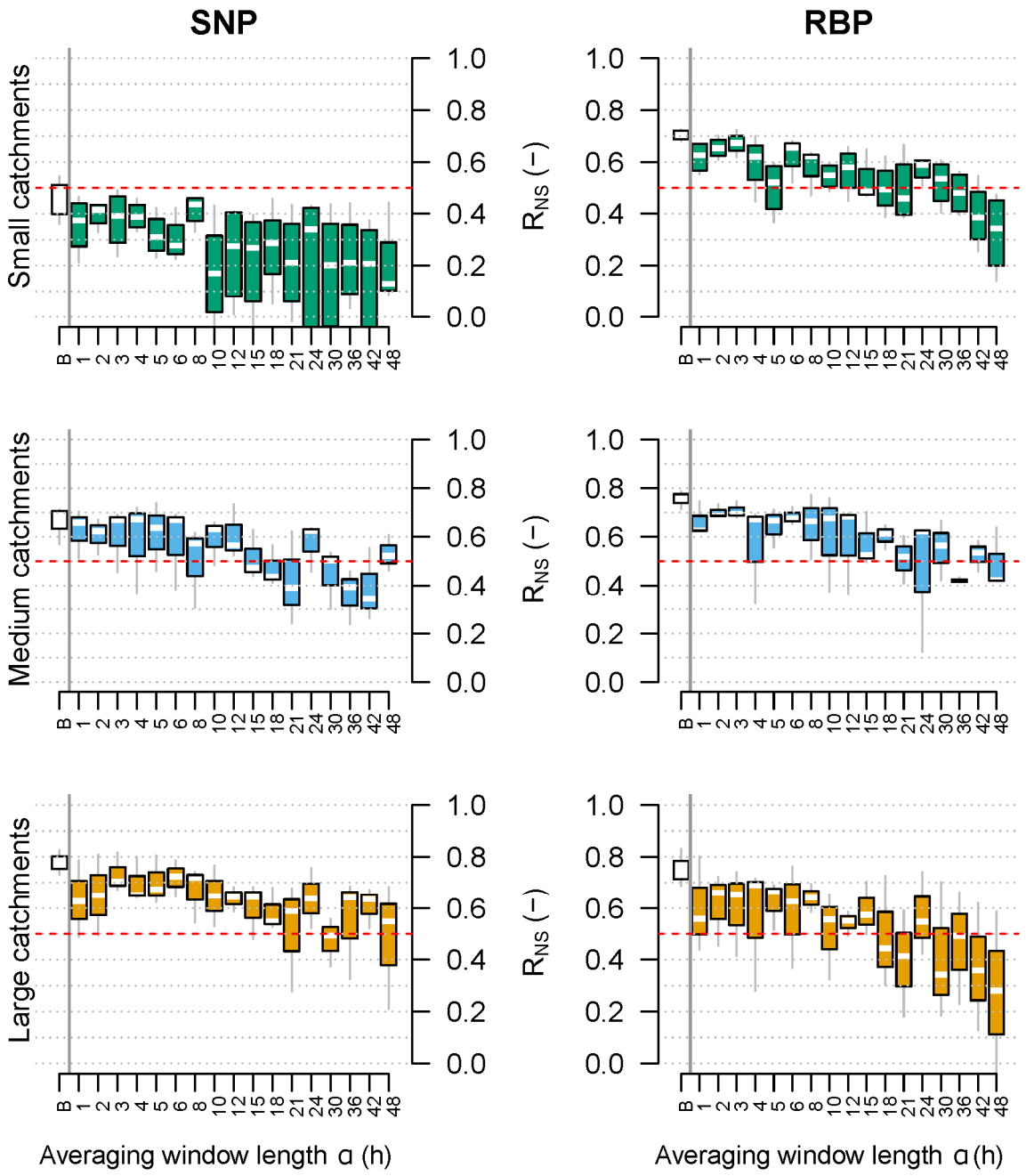

Figure 8. Averaged effect of the data temporal averaging (α) on model identification assessed by the Kling-Gupta efficiency with two precipitation datasets (RBP and SNP) grouped according to catchment area into small (A < 60 km^2); medium (60 ≤ A < 200 km^2); and large (A ≥ 200 km^2). The α labelled "B" corresponds to the benchmark.

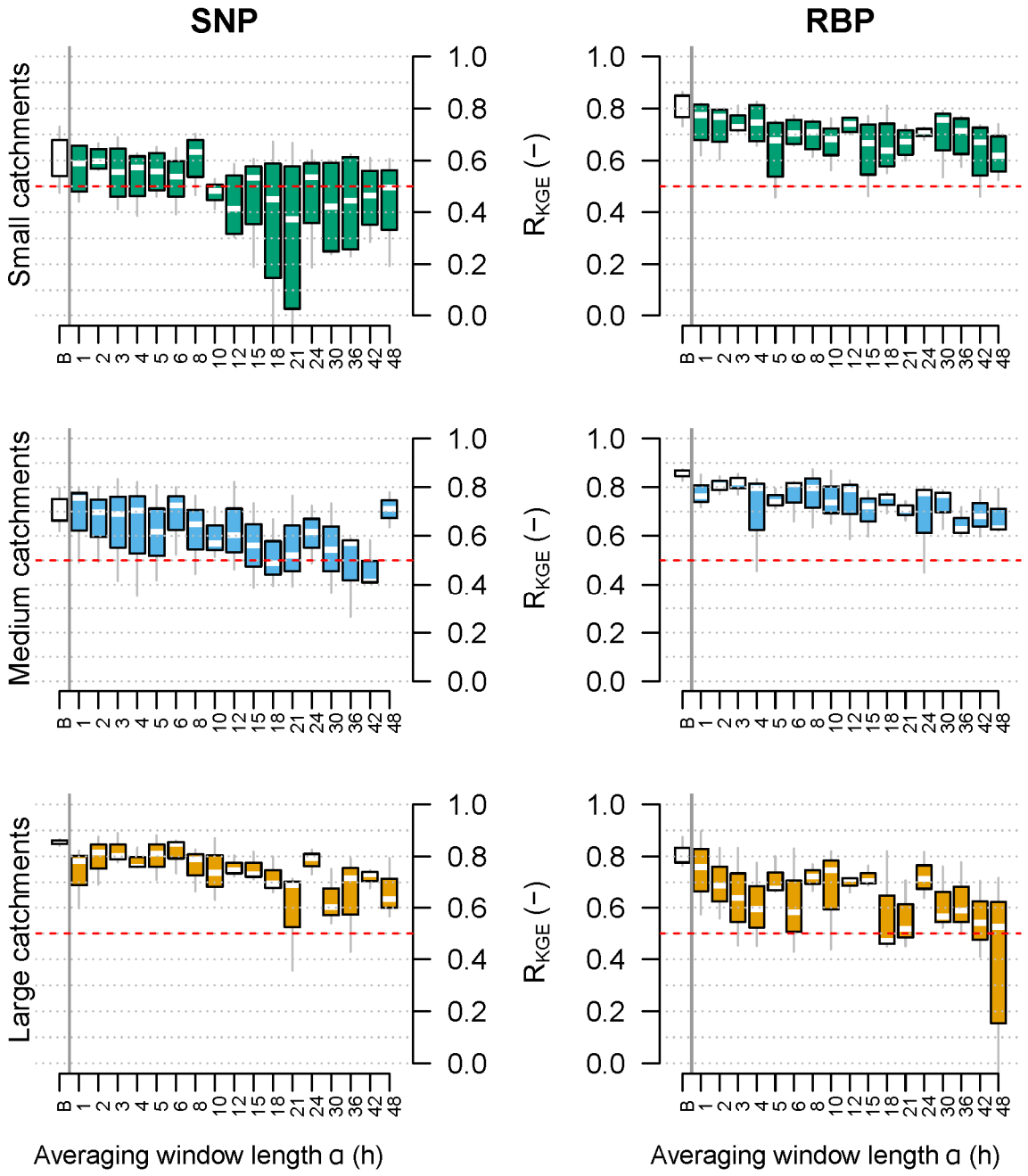

Figure 9. Estimated model error σ for 13 study catchments and two precipitation datasets (SNP and RBP) grouped by catchment size. Boxes depict the variability of σ within each size group and averaging window length. As no error is estimated in the benchmark approach, the B length here represents the mean prior value.

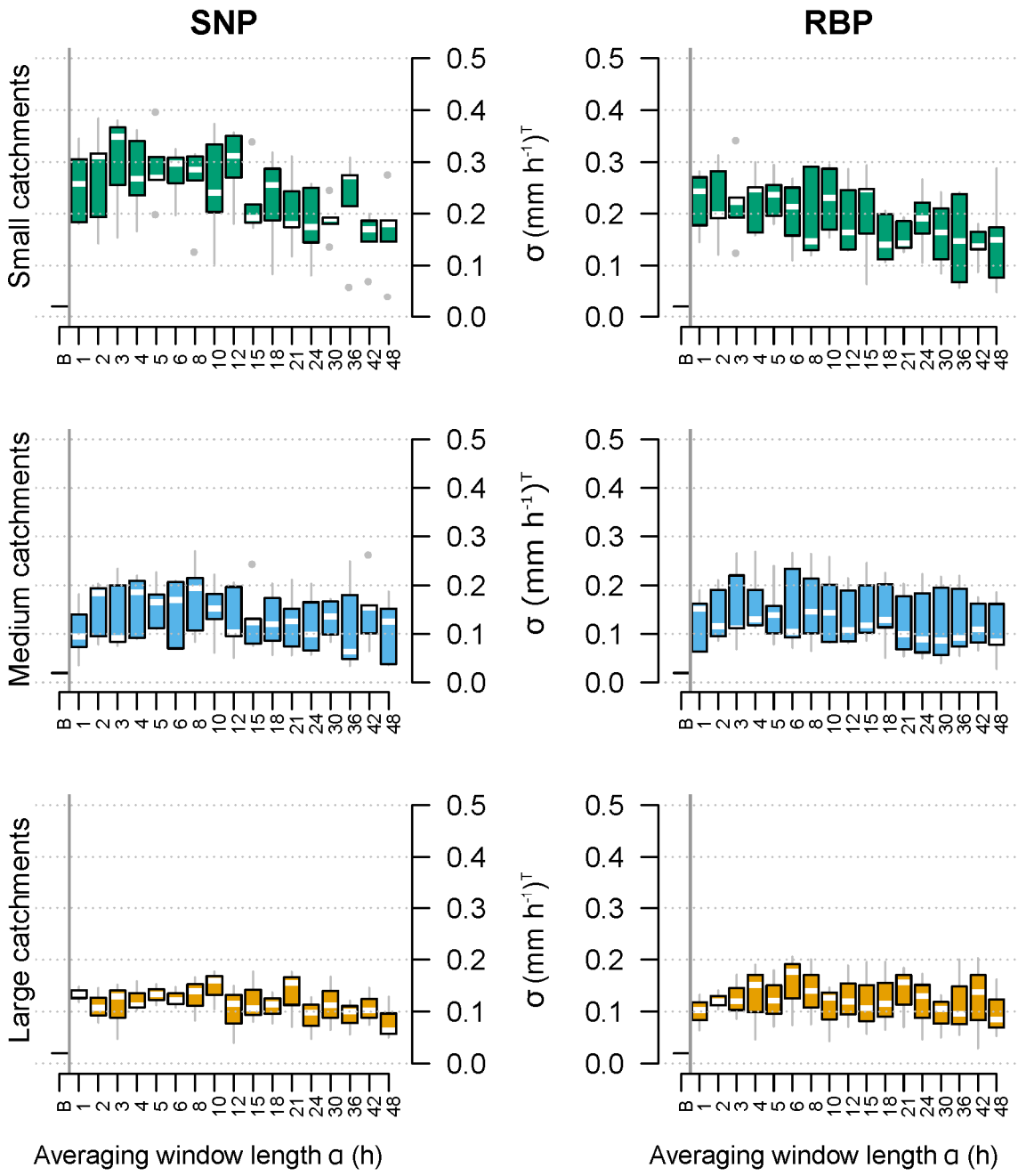

Figure 10. Sensitivity to corrupted SNP estimates *vs* uncorrupted SNP and RBP estimates assessed for different aggregation schemes based on three example catchments (C3, C8 and C12) and two efficiency metrics, R_{NS} (top) and R_{KGE} (bottom).